Detection of people and animals using non-imaging sensors

Thyagaraju Damarla
U.S. Army Research Laboratory
Adelphi, MD, U.S.A.
Email: thyagaraju.damarla@us.army.mil

Asif Mehmood
U.S. Army Research Laboratory
Adelphi, MD, U.S.A.
Email: asif.mehmood@us.army.mil

James Sabatier
U.S. Army Research Laboratory
Adelphi, MD, U.S.A.
Email: james.sabatier@us.army.mil

Abstract—**Non-imaging sensors offer low power and long lasting solutions for perimeter, border crossing, and forward operating base protection. In this paper, we study the utility of acoustic, seismic, and ultrasonic transducers for detection and identification of people and animals. Various algorithms will be developed for them, which are computationally less intensive and amenable to implement on sensor network. We identify the physics-based phenomenology associated with the targets and the features selected for classification are based on the phenomenology. We fuse the results from various sensor modalities to achieve higher probability of correct classification.**

Keywords: Personnel detection, sensor fusion, phenomenology, acoustic, seismic and ultrasonic.

I. INTRODUCTION

Personnel detection deals with the prevention, detection, and response to unauthorized persons from crossing an established perimeter [1]. It is required in a variety of military and civilian situations. Personnel detection is an important aspect of intelligence, surveillance, and reconnaissance (ISR). It plays a vital role in perimeter and camp protection and in curtailing illegal border crossings by people from neighboring countries, to name few [2] [3]. All these applications involve deployment of sensors for a prolonged time and often camouflaged to avoid discovery by others. Due to the low power requirement, the sensors used consist of non-imaging sensors such as acoustic, seismic, magnetic, E-field, passive infrared, ultrasonic, and radar. If imaging sensors are used, they are used to take a snapshot of the target to corroborate the findings by other modalities. In this paper, we consider a subset of the sensors listed above, namely, acoustic, seismic [4] [5] [8], and ultrasonic sensors [6] [7]. It will be clear throughout the paper that these three sensors are adequate to detect and identify people and distinguish them from other targets such as animals. However, no single sensor is adequate for the job. Fusion of the outputs or features from these sensors is the key for detection and classification with high confidence.

Detection and classification of any target should be approached via phenomenology of the target and sensor's ability to capture the phenomenology properly. This implies that the characteristics of the sensor should be adequate to capture the phenomenon being observed. For example, using a microphone with 1 kHz bandwidth will not do justice to music with 20 kHz bandwidth. Selection of the features for classification should represent the phenomenon being observed.

The main focus of this paper is to develop algorithms for detection of people, by understanding the underlying phenomenology of the signatures generated by humans and animals, and the detection of these signatures using multiple sensor modalities. Furthermore we process the data obtained by different non-imaging sensors to extract the phenomenology based features and apply algorithms to detect personnel.

This paper is organized as follows: Section II describes the data collection. Sensors modalities and target phenomenology are discussed in Section III. We also present various algorithms used to detect people in Section III and fusion of the results from multiple modalities. The paper is concluded in Section IV.

II. DATA COLLECTION

In order to develop algorithms based on real-world environments, we went to the Southwest border and collected data at three different locations, namely, (a) wash, a flash flood riverbed consisting of fine grain sand; (b) a trail, a trail formed by people walking through the thick of bushes and has the hard surface; and (c) choke point, a valley between two hills known to be trespassed by illegal aliens as shown in Figure 1. We used suite of sensors consisting of acoustic, seismic, passive infrared (PIR), magnetic & E-field, ultrasonic, profiling, radar sensors to collect the data. Some of the sensors used are shown in Figure 2. Each sensor suite is placed along the path with a spacing of 40 to 60 meters apart. Some of the scenarios used for data collection include: (a) a single person walking with and without back pack, (b) two people walking, (c) multiple people walking, (d) one person leading an animal, (e) two people leading animals, and (f) three people leading animals with and without payloads. A total of 26 scenarios with various combinations of people, animals, and payload are enacted and collected the data at those three sites. The data are collected over a period of four days; each day at a different site and different environment. Sometimes there is wind, sometimes it is quiet. The experiments with animals always involved people, hence, through out this paper animal detection using seismic and acoustic data analysis for cadence imply animal and person leading it.

Report Documentation Page

1. REPORT DATE **JUL 2011**	2. REPORT TYPE	3. DATES COVERED **00-00-2011 to 00-00-2011**

4. TITLE AND SUBTITLE **Detection of people and animals using non-imaging sensors**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **U.S. Army Research Laboratory,Adelphi,MD,20783-1197**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Presented at the 14th International Conference on Information Fusion held in Chicago, IL on 5-8 July 2011. Sponsored in part by Office of Naval Research and U.S. Army Research Laboratory.

14. ABSTRACT
Non-imaging sensors offer low power and long lasting solutions for perimeter, border crossing, and forward operating base protection. In this paper, we study the utility of acoustic, seismic, and ultrasonic transducers for detection and identification of people and animals. Various algorithms will be developed for them, which are computationally less intensive and amenable to implement on sensor network. We identify the physics-based phenomenology associated with the targets and the features selected for classification are based on the phenomenology. We fuse the results from various sensor modalities to achieve higher probability of correct classification.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**8**	

Figure 1. Different terrains: (a) Wash with fine grain of sand and (b) Trail
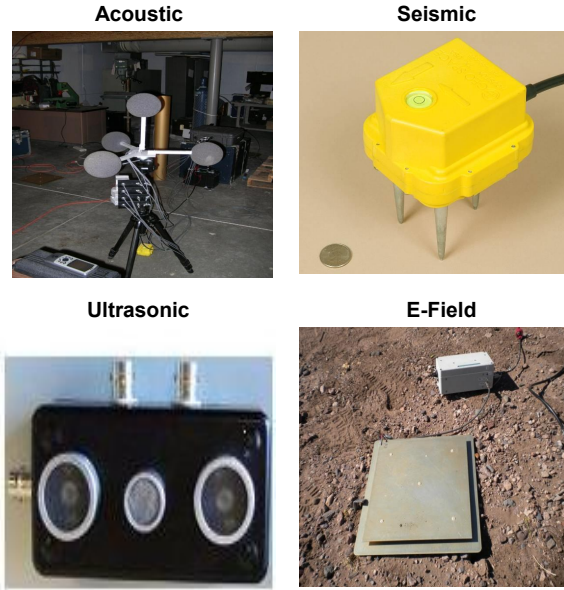
Figure 2. Acoustic, seismic, Ultrasonic and E-field sensors

Figure 3. Sample voice signal showing different words/consonants spoken

III. SENSOR MODALITIES, TARGET PHENOMENOLOGY, AND ALGORITHM DEVELOPMENT

In this section, we consider three sensor modalities shown in Figure 2, namely, (a) acoustic, (b) seismic, and (c) ultrasonic sensors for detection and classification of targets. As mentioned earlier, each sensor modality offers unique features that other modalities cannot. We present the target phenomenology associated with these modalities and the techniques used to exploit it, while keeping in mind that the these algorithms should be low complexity and amenable to implement on unattended ground sensors (UGS).
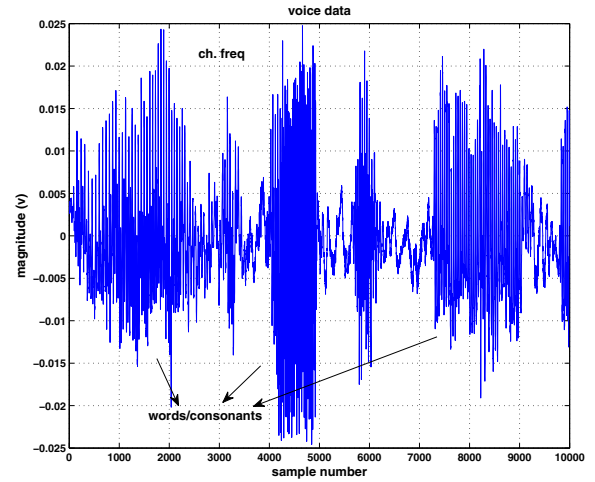
A. ACOUSTIC SENSOR DATA ANALYSIS

Humans depend heavily on hearing, next only to vision, to observe the targets and for better situational awareness. Humans also have the ability to perceive the targets without seeing by listening to the sounds the targets produce. In order to detect the presence of humans, we rely on the following phenomenological features extracted from:

- human voice and its characteristics
- sounds generated due to footfalls and their cadence

Human Voice: Humans generate sound by modulating the vocal cords and appropriately opening and closing the vocal tract [11]. In general, there are several frequencies associated with voice are called formants [11]. A small segment of a speech signal is shown in Figure 3. One would notice from Figure 3 that whenever a word is spoken a burst of high frequency signal appears and some background noise occurs during other times. This high frequency signal, the formant and varies from person to person and also depending on the word spoken. In general, the frequency lies between 200 - 800 Hz for the people we tested. Figure 4 shows the expanded version of the first segment of the voice signal shown in Figure 3 and Figure 4 shows its Fourier transform. Clearly, one can see the dominant frequency around 300 Hz. The objective of the signal processing is to detect and determine this frequency.

1) Detection of Personnel using Formants and Modulation Characteristics: As mentioned previously, the carrier frequency (formant) is amplitude modulated; its representation may be given as

$$s(t) = (A_c + A_m \sin \omega_m t)\cos \omega_c t \qquad (1)$$

where $\omega_c = 2\pi f_c$ and ω_m represent the carrier and modulating frequencies and A_c and A_m denote their magnitudes, respectively. The signal has three distinct frequency components, namely, f_c, $f_c + f_m$ and $f_c - f_m$. The spread of frequency (see Figure 4(b)) is then $\pm f_m$ around the carrier. The algorithm
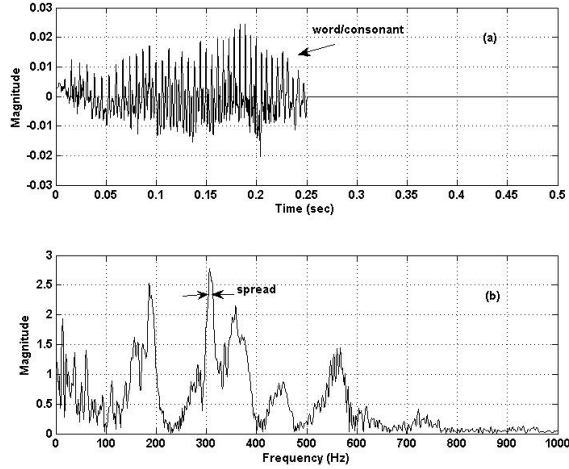
Figure 4. (a) Portion of voice signal in Figure 3, (b) its FFT

for detecting human voice consists of estimating the formant (carrier frequency) and the spread. If the spread is above some threshold, we declare it as a human voice. Statistical analysis is performed on various speech signals in order to determine the threshold value.

2) Personnel Detection using the Energy in Several Bands of Voice Spectra: It is known [11] that the human voice spans 50 Hz - 20 kHz frequency range. However, most of the energy is concentrated in 4 to 5 bands, as can be seen in Figure 4(b). These bands are 50 - 250 Hz, 251 - 500 Hz, 501 - 750 Hz, and 751 - 1000 Hz. The energy levels in these bands are the features and are designated by the feature vector $X = \{x_1, x_2, \cdots, x_n\}$, where x_i is the energy in band 'i', and n is the number of features. The feature vectors are used to classify whether they belong to human voice or not using a multivariate Gaussian (MVG) classifier as described in [2]. For the sake of continuity, we present a short description of the MVG classifier. We assume the energy levels in each band are statistically independent and have the Gaussian distribution given by

$$p(x_i) = \frac{1}{\sqrt{2\pi |\Sigma_i|}} \exp\left\{ -\frac{1}{2} (x_i - M_i)^T \Sigma_i^{-1} (x_i - M_i) \right\} \tag{2}$$

where M_i and Σ_i denote the mean and variance, respectively, and T denotes the transpose. Then the likelihood that a person is present or not is given by

$$p(X|H_j) = \Pi_{i=1}^n p(x_i|H_j) \, p(H_j), \quad j = \{0,1\} \tag{3}$$

where H_1 and H_0 are the hypothesis correspond to a person is present and not present, respectively. Then the posterior probability of human presence is given by

$$p(H_1|X) = \frac{\Pi_{i=1}^n p(x_i|H_1) \, p(H_1)}{\Pi_{i=1}^n p(x_i|H_1) \, p(H_1) + \Pi_{i=1}^n p(x_i|H_0) \, p(H_0)} \tag{4}$$

Assuming the priors $p(H_0) = p(H_1) = 0.5$, we can compute the posterior probability of a human present given X. If it exceeds a particular threshold value, we declare that a human is detected.

3) Personnel Detection using Cadence: Whenever a person or an animal walks, the footfalls make audible sounds. One can analyze the signatures of human and animal footfalls and classify them into respective classes. It is estimated that the cadence of the humans walking lies between 1 to 2 Hz while the cadence of animals walking is around 2.5 - 3 Hz. Moreover, these footfalls are impulsive in nature and result in several harmonics. Even if many people are walking in a file (on a path), they tend to synchronize their stride with others and walk more or less at the same cadence. This gives a way to estimate the cadence and then classify it. Cadence estimation and classification is similar to the algorithm for seismic data and is presented in the seismic data analysis section.
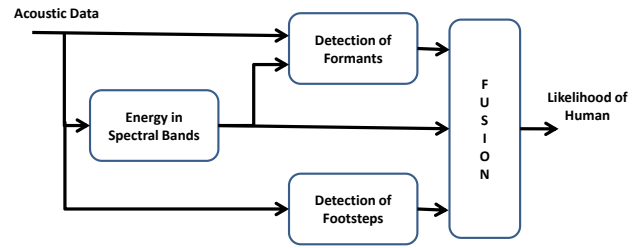

Figure 5. Acoustic data processing

Figure 5 gives the flowchart for processing acoustic data. The acoustic data are first analyzed to determine the presence of a person using the energy in spectral bands using MVG classifier. If the classifier gives the likelihood of a person greater than some threshold, the data are then further analyzed for the presence of formants. We also look for the presence of a person using cadence analysis. All three results are fused using Dempster-Shafer fusion paradigm [2], [12], and the results are shown in Figure 6. The top plot in Figure 6 is the original acoustic data collected in the field, the middle plot is the probability of detection of voice or footfall sound, and the bottom plot is the probability of detection of human voice by detecting formants. From the acoustic data plot we can see the impulses corresponding to the footfall sounds. The formant detection augments the fact that the sounds correspond to a person. The footstep detection using various harmonics of cadence is shown in Figure 7. The next section describes the seismic data analysis.

B. SEISMIC SENSOR DATA ANALYSIS

The main purpose of seismic sensors is to detect footfalls of humans walking within the receptive field of the sensor. There is a considerable amount of literature [1] - [10], [14] on footstep detection. Traditionally, estimation of cadence of the footsteps is performed for seismic data analysis. However, if multiple people are in the vicinity of the sensor and walking,
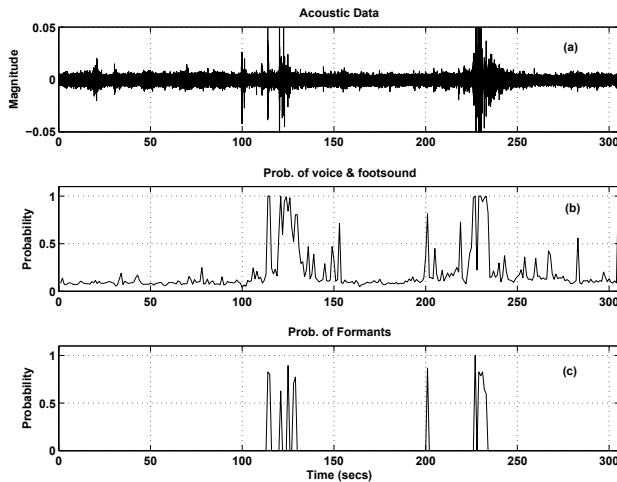
Figure 6. (a) Acoustic data of a person walking, (b) probability of voice/foot sound using MVG classifier and (c) probability of formant detection
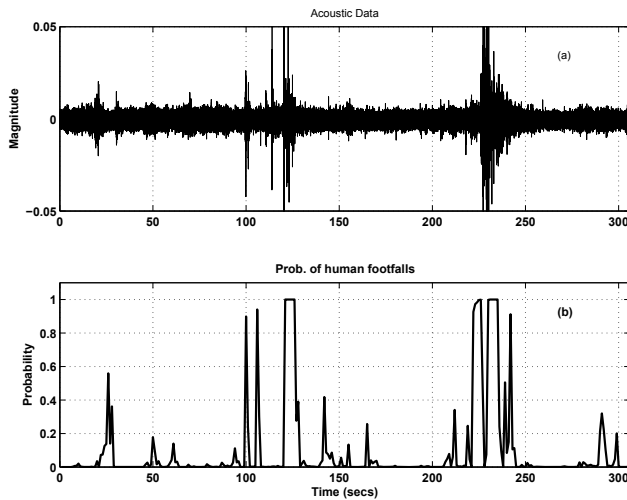

Figure 7. (a) Acoustic data of a person walking and (b) Probability of acoustic footstep detection
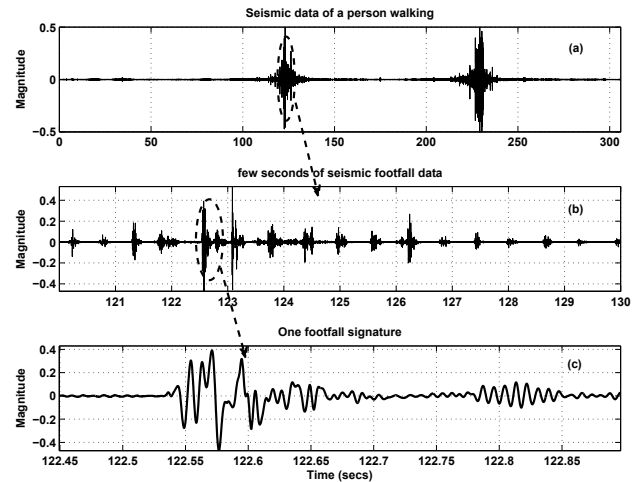

Figure 8. (a) Seismic data of a person walking, (b) enlarged portion – shows the periodicity of footsteps and (c) signature of one footstep
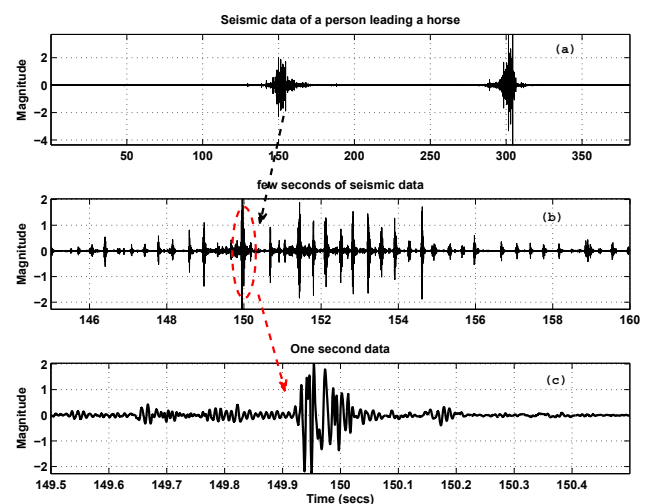

Figure 9. (a) Seismic data of a person leading a horse, (b) expanded portion – shows the periodicity of hoof signature and (c) enlargement of one impulse due to hoof

it is difficult to estimate the cadence of an individual person. Moreover, if there are animals, it is difficult to differentiate multiple people and animals walking by observing the footfalls. Figure 8 shows the signature of a person walking and Figure 9 shows the signature for a person leading a horse. However, the multiple footfalls superimpose one another, resulting in several harmonics of the cadence frequency 'c'.

To develop an algorithm for personnel detection with multiple people walking, jogging, running, or combination of them will be extremely difficult. In order to limit the scope of the problem, we assume that the people are walking on a path such as a paved road or trail in an open field. If there are animals, we assume that these animals are being led by people. We assume if people are running, they are running one behind the other with 3-4 m separation. Even though this restriction seems artificial, in fact, narrow trails form as

people walk and people tend to walk in single file as the trails are narrow; similarly, people use paved roads if they exist. If we assume that the people are walking on a path, the seismic signals due to footfalls of humans and animals exhibit a rhythm, and hence, has a cadence. When multiple people walk in single file they tend to synchronize their footsteps with one another for a majority of the time. Frequency analysis of the data would reveal the cadence of the person(s) or animal(s) walking. Since the seismic signals are impulsive in nature, several harmonics of cadence frequency can be observed in the frequency analysis. Since humans and animals have distinct cadences it is possible to classify the seismic signatures from them. We use the MVG classifier described earlier to do seismic signal classification. For the feature set, we first compute the spectrum of the envelope [1]–[3] of the

seismic signal accumulated for a period of 6 seconds. Then, the feature set $\{x_1, x_2, \cdots, x_n\}$ consists of amplitudes of the frequency bins from 2 to 15 Hz [2]. Then, the MVG algorithm is used to estimate the posterior probability of human or animal footsteps present. The results of the algorithm are shown in Figure 10.
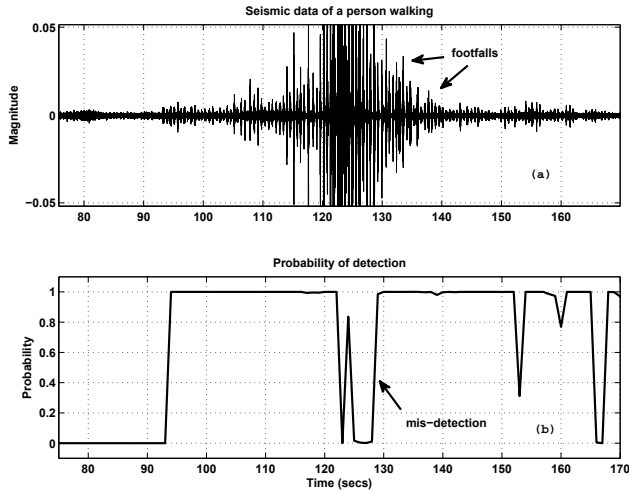

Figure 10. (s) Seismic data of a person walking and (b) probability of detection

The previously described classification works reasonably well if humans and animals are walking. However, if a person is running, the cadence of the person running is approximately the same as the cadence of a horse walking. In order to determine the presence of humans, it is necessary to determine whether these footsteps belong to a human or an animal. Additional signal processing is done to determine whether the seismic signatures belong to humans or animals. Figures 11 and 12 show some of the processing done on the signatures. Figure 11(a) shows the human footfalls and Figure 11(b) show the envelope of the magnitudes of the footfalls. The span is computed as the time duration when the magnitudes of the footfalls lie above some threshold. Similarly, Figure 12 shows the information for horse led by a person. Here we assume that the horse hoof signatures dominate the footfalls of a person leading it. The threshold is estimated to be the mean of the absolute values of the signatures. We use the magnitude of the signals along with the span of the signals above certain threshold as the features to determine the presence of humans or animals. Table I shows the features of a person walking and running and a horse walking. These features are used in a MVG classifier to classify the signatures.

1) Semantic Data Fusion: Seismic data are particularly sensitive to the soil conditions. Depending on the properties of the soil, the signals propagate at different velocities and the transfer function of the soil affects the signal differently. In order to perform the classification properly, it is necessary to use appropriate training set depending on the type of soil. The semantic tree used for classification is shown in Figure 13.
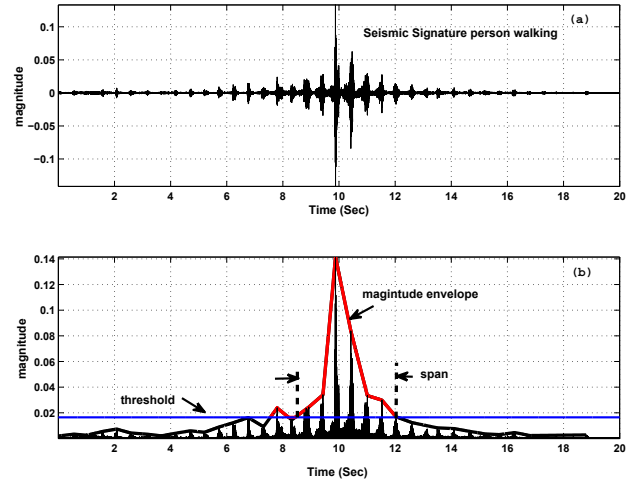

Figure 11. (a) Seismic signals generated by a person walking, and (b) signal span for a person walking
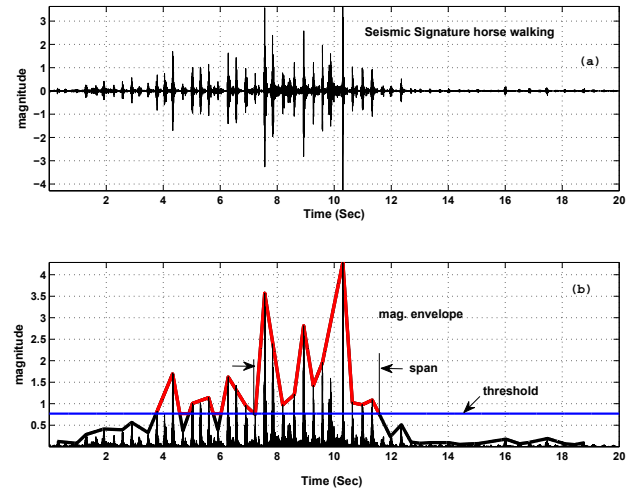

Figure 12. Seismic signals generated by horse led by a person, and (b) signal span

Table I
DISTINGUISHING FEATURES FOR PEOPLE AND ANIMALS

	Cadence	Peak Amplitude	Span
Person Walking	1.9 Hz	0.048	3.69 Sec
Person Running	2.79 Hz	1.21	3.34 Sec
Horse Walking	2.71 Hz	3.69	4.34 Sec

The semantic tree has two branches, namely, (a) wash and (b) trail, corresponding to two different soil conditions. The branch corresponding to the trail is expanded where the data are analyzed to determine the presence of personnel and animals. The branch corresponding to the personnel is analyzed to determine if the people are walking or running. Further analysis is done to determine if there is a single person
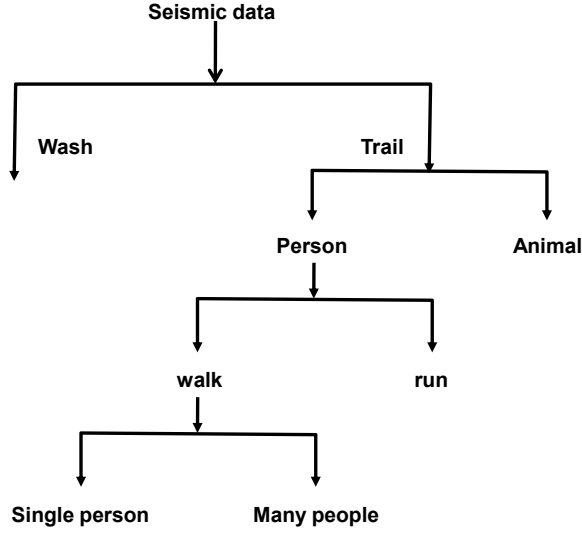
433

Figure 13. Semantic tree used for classification of seismic data

Figure 14. Micro Doppler from various body parts of a walking person.

Figure 15. Micro Doppler from various body parts of a walking horse.

or multiple people are present.

C. ULTRASONIC SENSOR DATA ANALYSIS

In this section, we discuss the processing of the ultrasonic data. The ultrasonic data are rich in information and embody the Doppler signature of a moving human or an animal such as a horse [6]. Typical Doppler velocities that are proportional to the Doppler frequencies from various body parts of a walking human and from a walking horse are shown in Figures 14 and 15, respectively . Ideally, the Doppler from the arm, leg and torso of a person is different from that of animal legs. As mentioned previously, it is important to know the number of people and animals to perform classification. This is due to the reason that information about the number of people and animals has to be included in the training data set. Towards this goal, we processed the ultrasonic data to count the number of targets in the vicinity using the energy content in various bands of Doppler. Figure 16 shows the flowchart for the algorithm used in counting the number of targets. For processing the ultrasonic data a 1 second interval of the data is considered at a time and the algorithm shown in Figure 16 is used to find the energy in each band. Then a sliding window is used, which slides approximately 0.1 second and next segment of data is obtained and processed. The algorithm results for several runs are shown in Figure 17. The scenarios used corresponds to (a) one man walking, (b) one man leading an animal, (c) two men and one woman walking and, (d) four men and three women walking. In the last case, a count of only six targets are realized using the algorithm. The reason is due to a large number of people, one is very close to the other, masking the Doppler returns from one.

1) Classification of targets using ultrasonic data: The Doppler returns from animals are quite different compared to those from 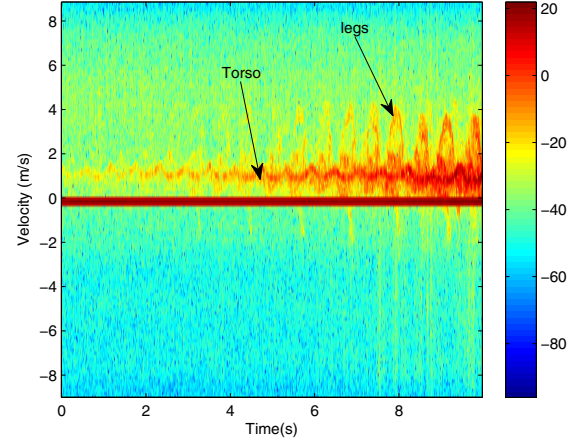humans. One distinction is that humans have stronger returns from their torsos while animals have significantly weaker Doppler returns from their torsos, as is evident from Figures 14 and 15. The total energy in various bands for the animal is different from that of the humans, as shown in Figure 17. In order to classify, 40 features are selected from each band B_i, $i \in \{1, 2, 3\}$
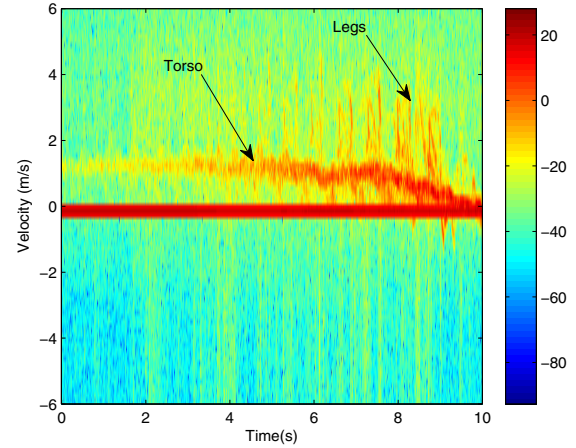
$$\mathcal{F}^{B_1} = \left\{ F_1^{B_1}, F_2^{B_1}, \cdots, F_{40}^{B_1} \right\}$$

where $F_k^{B_i} = \frac{1}{5} \sum_j^{j+4} f_j$ where $j = (k-1)*5+1+C_i$, f_j is the magnitude of the Fourier coefficient j, and $C_i = \{100, 300, 500\}$ for the band B_i. Training data are generated for each point on Figure 17 that corresponds to people, animal, and everything else. There are three classes, namely, (a) human, (b) animal, and (c) others. We developed a support vector machine with a Gaussian kernel to perform classification. The correct classification of 95% are achieved. When we used only two classes, humans and everything else (that is, animal plus others), we achieved a correct classification of 98%.
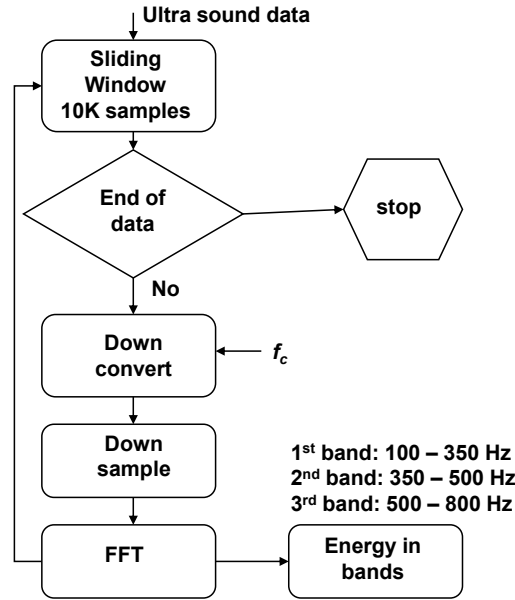
434

Figure 16. Flowchart showing the ultrasonic signal processing for counting number of targets

1st band: 100 – 350 Hz
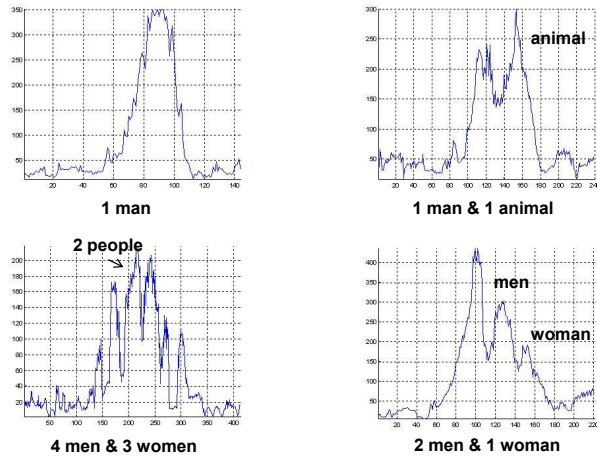2nd band: 350 – 500 Hz
3rd band: 500 – 800 Hz

Figure 17. Target count using ultrasonic data analysis

D. COMPLETE IMPLEMENTATION OF PERSONNEL DETECTION ALGORITHM

The previous sections showed how each individual sensor modality data is processed to detect and classify personnel. We determined that in order to get better classification with fewer false alarms, it is necessary to know the number of targets in the sensor receptive area as well as to use the right training data for classification depending on the type of site, for example, the wash, trail, etc. Figure 18 shows the tree structure used to detect personnel.

In the hierarchical structure, we first use the ultrasonic data analysis to determine the number of targets present in the
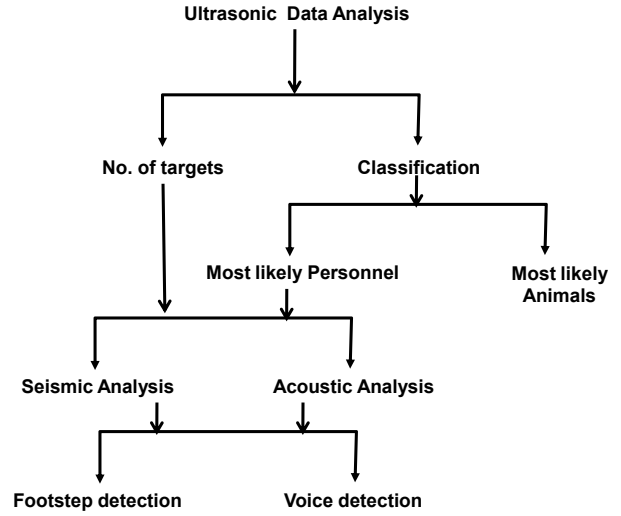

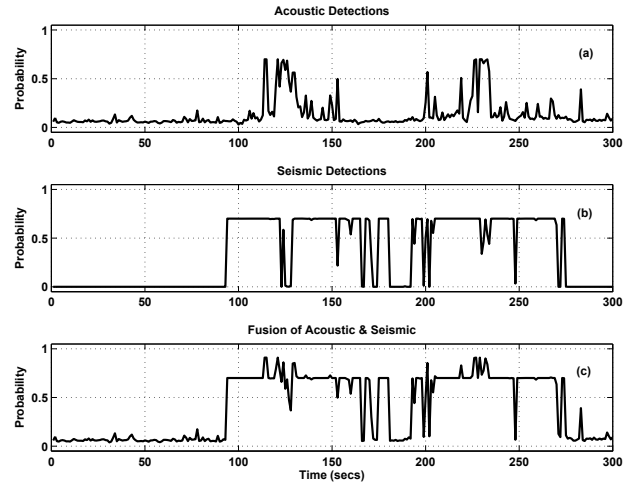Figure 18. Hierarchical structure used for personnel detection

Figure 19. (a) Detection of human voice using acoustic sensor, (b) Detection of footsteps by seismic sensor, and (c) Fusion of acoustic and seismic information

vicinity of the sensor field and then determine the likelihood of people present. If it is determined that there is high likelihood of people present, then we use both acoustic and seismic data to further corroborate the presence of people.

The acoustic and seismic sensors used for collection were co-located while the ultrasonic sensor is located about 20 meters away from the acoustic and seismic sensors. Moreover the ultrasonic sensor data is not time synchronized with the others. As a result, we can not fuse the information from all three. However, we can determine the presence of people and animals using the ultrasonic data. Once, the presence of people is established, the acoustic and seismic data is fused and the results are shown in Figure 19. Fusion is accomplished using Dempster-Shafer fusion [2], [12], [13] paradigm. The uncer-

435

tainty of each sensor is established based on the classification of data used for training. The uncertainty for both acoustic and seismic data is found to be 30%. As a result the probability of detection values for either acoustic or seismic data does not exceed 0.7 as can be seen in Figure 19(a) and (b). However, the fusion of acoustic and seismic information resulted in higher probability of detection Figure 19(c).

IV. Conclusions

In this paper, we presented several algorithms for personnel detection using acoustic, seismic, and ultrasonic data. The acoustic data are analyzed for formants and footstep detection. The acoustic data are also used to estimate the cadence of animals walking and discriminate between animals and people when a human voice is not present. Seismic data are analyzed for footstep detection and classification of humans and animals. We used ultrasonic data for estimating the number of targets present and for classification. We were able to achieve high percentage of correct classification using all three sensor modalities. The complete suite of algorithms with other modalities is still being developed and will be evaluated for false alarms. Each algorithm tried to use the sensor's particular phenomenology for the detection and classification of people. The algorithms presented are computationally efficient, consume less power and hence amenable for implementing on sensor networks such as networked UGS.

References

[1] K. M. Houston and D. P. McGaffigan,"Spectrum Analysis Techniques for Personnel Detection Using Seismic Sensors", *Proc. of SPIE*, Vol. 5090, pp. 162-173, 2003.

[2] Raju Damarla and David Ufford, "Personnel detection using ground sensors", *Proc. of SPIE*, Vol. 6562, 656205, 2007.

[3] Thyagaraju Damarla, James Sabatier, and Alex Ekimov, "Personnel detection at a border crossing", Proc. of Military Sensing Symposium National, July 2010, Las Vegas.

[4] G. Succi, D. Clapp and R. Gambert, "Footstep, Detection and Tracking", *Proc. of SPIE*, Vol. 4393, pp. 22-29, 2001.

[5] A. Sunderesan, A. Subramanian, P. K. Varshney and T. Damarla, "A copula based semi-parametric approach for footstep detection using seismic sensor networks", *Proc. of SPIE*, Vol. 7710, 77100C, 2010.

[6] A. Mehmood, J. M. Sabatier, M. Bradley and A. Ekimov, "Extraction of the velocity of walking human's body segments using ultrasonic Doppler", *The Journal of the Acoustical Society of America*, pp. EL316-EL322, vol. 128, no.5, Oct. 2010.

[7] Alexander Ekimov and James M. Sabatier, "Passive ultrasonic method for human footstep detection", *Proc. of SPIE*, Vol. 6562, 656203, 2007.

[8] James M. Sabatier and Alexander Ekimov, "Range limitation for seismic footstep detection", *Proc. of SPIE*, Vol. 6963, 69630V-1, 2008.

[9] R. E. Bland, "Acoustic and seismic signal processing for footstep detection", *Master's Thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science*, 2006.

[10] S. G. Iyengar, P.K. Varshney and T. Damarla, "On the detection of footsteps based on acoustic and seismic sensing", *Conference Record of the Forty First Asilomar Conference on Signals, Systems and Computers, ACSSC 2007*, pp. 2248-2252, 4-7 Nov. 2007.

[11] Encyclopedia of Acoustics, Vol. 4, Edited by Malcolm J. Crocker, Published by John Wiley & Sons, Inc. New York, NY., 1997.

[12] H. B. Mitchell, An Introduction to Multi-Sensor Data Fusion, Springer-Verlag, New York, 2007.

[13] D. L. Hall and S. A. H. McMullen, Mathematical Techniques in Multisensor Data Fusion, Artech House, Norwood, MA, USA, 2004.

[14] H. Park, A. Dibazar and T. Berger, "Cadence analysis of temporal gait patterns for seismic discrimination between human and quadruped footsteps", IEEE Conference on Acoustics, Speech and Signal Processing, pp. 1749-1752, 2009.